利用 PSO-SA 混合优化支持向量回归的径流预报模型研究 *

蒋林利 [1]，李　洁 [1]，吴建生 [1,2]

(1. 广西科技师范学院 数学与计算机科学学院，广西 来宾 546199; 2. 武汉理工大学 信息工程学院，武汉 430070)

摘　要： 为了有效提高径流预报的准确度，提出一种有效的融合优化策略，采用基于粒子群和模拟退火算法相结合的混合方法同时优化支持向量回归核函数类型和内核参数，以此建立一种有效的混合优化支持向量回归径流预报模型。提出的方法为核函数选择和参数优化提供了一种有效途径。通过对广西柳州柳江径流实例分析，并与纯粹的支持向量回归模型对比，研究结果表明，该模型预测稳定，具有较高泛化性能和预测准确度，为径流预报提供了一种有效预测方法。

关键词： 支持向量回归；粒子群；模拟退火；融合改进；径流预报模型

中图分类号： TP301.6　　doi: 10.3969/j.issn.1001-3695.2018.03.0148

Runoff forecasting based on hybrid optimized support vector regression using PSO-SA

Jiang Linli[1], Li Jie[1], Wu Jiansheng[1, 2]

(1. *Dept of Mathematics & Computer Science Guangxi Science & Technology Teachers College*, *Laibin Guangxi* 545004, *China*; 2. *School of Information Engineering*, *Wuhan University of Technology*, *Wuhan* 430070, *China*)

Abstract: In order to effectively improve the accuracy of runoff forecasting, this paper proposed an effective hybrid optimization strategy which was based on the combination of particle swarm optimization and simulated annealing algorithm, and it also optimized the type of kernel function and the kernel parameter setting of Support Vector Regression to establish an effective hybrid optimization support vector regression runoff forecasting model. The proposed method provided an effective way for the choice of kernel functions and parameter optimization. By analyzing the examples of Guangxi Liujiang's River runoff and with pure support vector regression model comparison, the results of the study show that the model is stable in prediction and it has high generalization performance and accuracy of prediction, and it can provide an effective prediction method for runoff forecast.

Key words: support vector regression; particle swarm optimization; simulated annealing; fusion improvement; runoff forecast model

0　引言

准确及时的降雨径流预报是水资源管理一个非常重要的课题，因为有效准确的降雨径流预报有助于指导提前防备，可以有效避免如生命安全、经济损失等的意外事故[1]。然而，由于存在影响水文的各种非线性因素[2]，如降雨特征、流域形态、水位和土壤含水量等，一个相对稳固、准确的降雨径流模型的开发已经成为了水文学领域中一个非常困难的课题[3]。此外，数据的缺失、噪声数据和无效数据更是加重了降雨径流模型研究中数据的匮乏[4]，模型构建面临更大的挑战。

神经网络(NNS)[5]能够在潜在关系未知或者难以描述的情况下，发现数据之间微妙的函数关系，已经成为模拟降雨径流预报一个非常有用的方法[6]。然而，神经网络也存在着一些不足，如一系列控制参数的获取，难以获得一个稳定的解决方案和存在模型过拟合的风险。

近年来，支持向量机(support vector machine，SVM)得到不断发展，这是一种基于统计学习理论的学习机，坚持结构风险最小化原则，并坚持寻求最大限度地减少所产生误差的上限，而不是训练误差最小化[7]。随着 Vapnik 对不敏感损失函数的引入，支持向量机的回归模型，即支持向量回归(support vector regression，SVR)，在解决非线性估计问题上也得到了越来越多的关注。SVR 成为了解决大规模数据集分类问题的有力方法。并且由于其许多有吸引力的特点和良好的泛化性能，在学术界和工业界平台上已经取得了很大的成功。

然而，使用支持向量机的回归时都面临两个问题：如何选择 SVR 的核函数类型以及如何确定 SVR 核函数的最佳参数，

收稿日期： 2018-03-01；**修回日期：** 2018-04-13　**基金项目：** 国家科技部中小企业创新基金项目（13C26214504766）；广西科技厅面上基金资助项目（2014GXNSFAA118027）；2017 年广西高校中青年教师基础能力提升项目（2017KY0896）；广西教育厅高校科研项目（YB2014467）

作者简介： 蒋林利（1976-），女，广西全州人，副教授，硕士，主要研究方向为机器学习、数据挖掘等；李洁（1980-），女（通信作者），广西贵港人，副教授，硕士，主要研究方向为神经网络（lijie980522@163.com）；吴建生（1974-），男，陕西咸阳人，教授，硕士，主要研究方向为人工智能神经网络.

这两个问题至关重要。因为核函数类型影响支持向量回归参数，参数的优化又决定了核函数的性能，两者相互影响和作用，需要协同考虑。SVR 的性能高度依赖于 SVR 的核函数类型和 SVR 核函数的参数设置。恰当的 SVR 核函数类型和 SVR 核函数的参数可以提高 SVR 回归精度。而不当的 SVR 参数将会导致过拟合或者低拟合[8]。普通的 SVR 模型中，支持向量回归模型产生的效果很大程度上依赖于操作者的经验，操作人员依据经验通过试凑获得最优且有效的参数，然而稍不小心，便容易使模型陷入过拟合，这样的模型可能会很好地预测过去的事件，但无法预测未来事件。

与此同时，基于智能群体的粒子群优化算法(particle swarm optimization，PSO)已经被广泛地应用于优化核函数参数。Kennedy J 于 1995 年提出的粒子群优化算法收敛速度快，能不断更新粒子的位置及速度来搜索最优解。对一些相对简单的问题，简单的优化算法便可以得到满意的结果。但是对于多维的非线性优化问题，粒子群优化算法在优化过程中容易出现局部收敛现象，从而导致陷入局部最优解。为了使粒子群算法避免局部最优解，研究人员提出了各种方法弥补 PSO 缺陷，改进 PSO 算法优化模型。吴建生等提出使用 BP 算法避免粒子群优化算法在局部搜索过程中的收敛停滞现象[9]。而模拟退火算法（Simulated Annealing，SA）是一种良好可有效避免陷入局部极小值，并能够在全局寻找最优解的优化算法，可以有效的弥补 PSO 算法的缺陷。

近年来的一些研究也已经相继提出了针对高斯核函数参数优化的算法，如遗传算法(genetic algorithm，GA)、粒子群优化(PSO)和免疫算法(immune algorithm，IA)[10~12]，但没有出现针对 SVR 核函数类型的选择算法。

为此，本文提出一种模拟退火和粒子群混合算法优化 SVR 的核函数类型及其参数，命名为 SVR-HPSOSA。该算法能同时确定 SVR 模型适当的核函数类型和最优核参数。SVR-HPSOSA 算法可以使粒子群优化算法避免过早陷入局部最优解，能够找到全局最优解。并应用于广西柳州市的柳江径流数据中，建立基于混合粒子群及模拟退火算法进化支持向量回归的径流预报模型。结果表明，运用 SVR-HPSOSA 模型进行日降雨量径流预报，显示了较高的模型预测精度和泛化能力。

1　支持向量回归

1.1　支持向量回归模型

支持向量机回归分析的主要思想是：假设 $(x_i,d_i)_{i=1}^{N}$ 是给定的训练数据，x_i 是输入向量，d_i 是输出向量，N 为数据的总维数。线型回归函数定义如下式：

$$f(x)=\omega\phi(x_i)+b \tag{1}$$

其中：ω 和 b 是相关系数，$\phi(x)$ 是从输入空间到高维特征空间的非线性映射。

支持向量机的目的是在不敏感函数训练模式中保证 $\|\omega\|^2$ 规范尽可能的小。根据文献[13],支持向量机函数为

$$f(x,\alpha_i,\alpha_i^*)=\sum_{i=1}^{N}(\alpha_i,\alpha_i^*)K(x,x_i)+b \tag{2}$$

其中：$K(x,x_i)$ 为核函数。在机器学习理论中，有如下四种常用的核函数：

$$\text{线性核函数}: K(x_i,x_j)=x_i^T\cdot x_j \tag{3}$$

$$\text{多项式核函数}: K(x_i,x_j)=(t+x_i\cdot x_j)^d \tag{4}$$

$$\text{高斯核(RBF)函数}: K(x_i,x_j)=\exp(-\frac{\|x_i-x_j\|^2}{2\sigma^2}) \tag{5}$$

$$\text{S 曲线型核函数}: K(x_i,x_j)=\tanh[\alpha(x_i^T\cdot x_j)+\beta] \tag{6}$$

1.2　支持向量回归参数

支持向量回归方法是近年来出现的解决非线型回归问题的一种高效方法。设计一个有效的支持向量回归模型，需要精心挑选支持向量回归的四个基本参数。不同的参数设置会对支持向量机性能产生显著的差异。这要确定的四个参数包括：

a）核函数类型。核函数用于在输入空间和特征空间之间构造一个非线性映射。

b）内核参数。包括表示多项式核的拦截参数 t，代表多项式核的程度系数 d，表示高斯核函数的带宽 σ，以及 sigmoid 核函数的参数 α 和 β。

c）正则化参数 C。代表最大限度地减少训练误差和减少模型复杂度的产生成本。

d）ε-不敏感损失函数。它等效于在训练数据点的近似精度。

2　HPSOSA 混合优化支持向量回归策略

2.1　粒子群优化算法

PSO 是由 Kennedy 和 Eberhart 在 1995 发明的一种进化计算技术，它已经被成功地用于解决复杂的优化问题，由于其在简单仿真系统中的灵活性和有效性，引起了研究者的兴趣。粒子群优化算法源于对鸟群捕食的行为研究。粒子群优化算法的基本思想是通过群体中个体之间的协作和信息共享来寻找最优解。

PSO 的优势在于简单容易实现并且没有许多参数的调节。目前已被广泛应用于函数优化、神经网络训练、模糊系统控制以及其他遗传算法的应用领域。但是近年的研究已经指出了其在应用领域的不足，如由于过早收敛导致在搜索过程中发生粒子群最优解陷入局部最小的缺点。

粒子群优化算法的算法思想是：粒子 i 在 N 维空间的位置表示为矢量 $X_i=(x_1,x_2,\cdots,x_N)$，飞行速度表示为矢量 $V_i=(v_1,v_2,\cdots,v_N)$。每个粒子都有一个由目标函数决定的适应值(fitness value)，并且知道自己到目前为止发现的最好位置(P_{best})和现在的位置 X_i。这个可以看作是粒子自己的飞行经验。

除此之外，每个粒子还知道到目前为止整个群体中所有粒子发现的最好位置(G_{best})(G_{best} 是 P_{best} 中的最好值)。这是粒子同伴的经验。粒子就是通过自己的经验和同伴中最好的经验来决定下一步的运动。

PSO 初始化为一群随机粒子(随机解)。然后通过迭代找到最优解。在每一次的迭代中，粒子通过跟踪两个“极值”(P_{best} , G_{best})来更新自己。找到这两个最优值后，标准 PSO 算法粒子通过下面的公式来更新自己的速度和位置。

$$V_i = \omega \times V_i + c_1 \times rand() \times (Pbest_i - x_i) + c_2 \times rand() \times (Gbest_i - x_i) \quad (7)$$

$$x_i = x_i + V_i \quad (8)$$

上面两式中 $i = 1, 2, \cdots, M$ ，M 是该群体中粒子的总数；ω 非负，称为惯性因子；V_i 是粒子的速度；P_{best} 和 G_{best} 如前定义；$rand()$ 是介于 (0,1) 之间的随机数；X_i 是粒子的当前位置。$c1$ 和 $c2$ 是学习因子，通常取 $c1 = c2 = 2$；在每一维，粒子都有一个最大限制速度 $V_{\max}$ ，如果某一维的速度超过设定的 $V_{\max}$ ，那么这一维的速度就被限定为 $V_{\max}$ （ $V_{\max} > 0$ ）。

2.2　模拟退火算法

模拟退火算法（SA）是一种随机全局搜索优化算法，它是基于统计热力学来模拟退火行为过程[14]。可以将退火过程中系统熵值类比为目标函数值 F，来模拟这个退火系统的算法。1983 年，Kirkpatrick 等提出模拟退火算法，并将其应用于组合优化问题的求解。模拟退火算法是一种可以有效避免陷入局部极小并最终趋于全局最优的串行结构的优化算法。

算法遵循 Metropolis 抽样稳定准则，即以概率接受新状态。假设在状态 x_{old} 时，系统受到某种扰动而使其状态变为 x_{new}。与此相对应，系统的能量也从 $E(x_{old})$变成 $E(x_{new})$，系统由状态 x_{old} 变为状态 x_{new} 的接受概率 p 定义为

$$p = \exp[-(E_j \cdot E_i) / k_B T] \quad (9)$$

模拟退火算法的实现流程如图 1 所示。

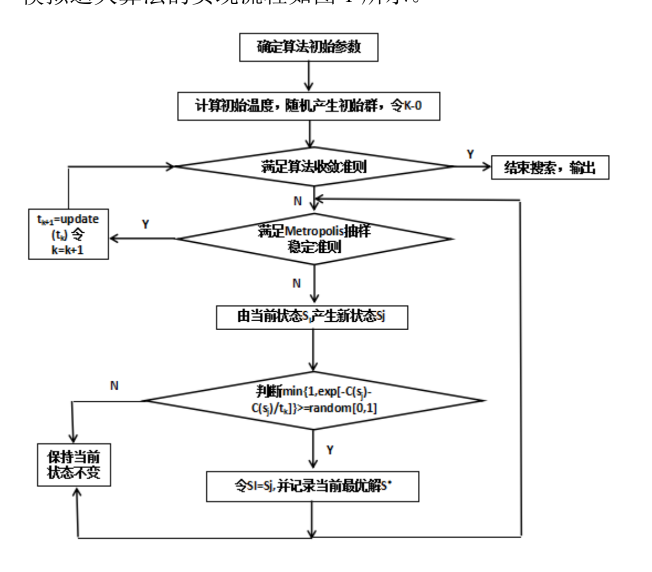


图 1　SA 算法流程图

2.3　HPSOSA 混合优化支持向量回归参数

为了得到全局或者接近全局的最优解，本文提出了一种混合粒子群和模拟退火算法的混合优化算法(HPSOSA)，充分利用了模拟退火算法能避免局部最优解获得全局最优解及粒子群算法速度快容易执行的优势。HPSOSA 算法容易实现，综合了粒子群和模拟退火算法两者的优点。

一般说来，当基于一个小而且适当的特征子集选择参数时，大多数的研究人员仍然遵循试验-误差的方式，首先，基于不同的参数集构建几个 SVR 模型，在验证集上进行测试，从而获得最优参数。然而，这个过程是耗时的，而且需要一些运气。除了适当的核函数类型外，参数的优化设置也可以提高 SVR 模型的准确度。而核函数类型会影响核参数，为了得到全局最优的 SVR 核函数和最优的 SVR 参数，两者应同时考虑。

本文采用 HPSOSA 算法同时优化核函数类型和核参数设置，建立基于 HPSOSA 优化的支持向量机回归模型，将改进的 SVR-HPSOSA 模型应用于中国广西柳州市柳江的日降雨径流预报中。

2.4　染色体表示

表1概括出了需要优化的各种核函数类型、有效的核函数参数以及SVR参数。因此，将染色体设计成三部分组成。核函数类型（整数），内核参数以及SVR参数（实数）。表1显示出了具有 d 维染色体的表示，其中 d 是不同数据集变化的特征数。染色体00,01,10,11表示核函数类型，这四个值分别表示系统将选择线性核函数，多项式核函数，高斯核函数和S型核函数。染色体 $\{x_{i,1}, x_{i,2}\}$ 代表了内核参数的值，染色体 $\{x_{i,3}, x_{i,4}\}$ 分别代表了惩罚参数及不敏感损失函数的值。

表 1　核函数类型、参数及支持向量回归参数表

核类型	核参数		SVR参数		整型编码	实型参数编码
线性核	γ	-	C	ε	00	$x_{i,1}, x_{i,2}, x_{i,3}$
多项式核	d	d	C	ε	01	$x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}$
高斯核	C	σ	C	ε	10	$x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}$
S曲线形核	α	β	C	ε	11	$x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}$

说明：γ,d,t,C,σ,β和α表示核函数参数变量

2.5　SVR–HPSOSA 模型

图2说描述了支持向量回归优化算法SVR–HPSOSA的过程。本文提出的SVR–HPSOSA算法具体描述过程如下：

a)粒子的初始化及粒子群参数设置：生成SVR内核类型及参数的初始粒子组合。设 $k = 0$ ，并随机初始化粒子群位置 s_i^0 以及粒子群的速度 v_i^0 。

b)输入训练数据并根据式子(11)计算每个染色体的适应值。

c)评估所有粒子的适应价值，通过对适应值的简单比较确定 G_{best} 和 P_{best} 。适应函数被定义为对训练数据集采用 $MAPE_{crossv-alidation}$ 函数的5次迭代交叉验证，表示如下：

$$M\inf = MAPE_{crossv-alidation} \quad (10)$$

$$MAPE_{crossv-alidation} = \frac{\sum_{i=1}^{l} |a_i - p_i| / a_i}{l} \times 100\% \quad (11)$$

其中：l 是训练数据样本数；a_i 是实际值，p_i 是预测值。在本研究中，5次迭代交叉验证法提供了计算成本和有效参数估计值之间的最佳折衷，被用于在训练数据集上计算适应值。对每一个粒子，通过式(11)计算他的适应值。

d)执行粒子群算法。根据适应度评估结果更新全局和个体值。设置平均训练变化精度为全局值，验证精度为个体最佳值。

e)计算 $\Delta Fitness = Fitness(s_i^{k+1}) - Fitness(s_i^k)$，随机选择一个数 $U \in [0,1]$。如果 $\Delta Fitness > 0$，意味着适应度函数增加，新的位置发生变化。新位置 s_i^{k+1} 根据的如下原则设置：$\exp(\frac{\Delta Fitness}{T_i}) > U$，当所有粒子的速度都已经确定，转向第f)步，否则，转向步b)，用相同的估计方法为尚未被接受的粒子生成新的速度。在我们研究的100个例子中，相同的粒子将迫使最后的速度被接受，同时考虑机器的执行时间，可以防止进入无休止的循环。对于可能的失败并没有发生在我们的研究和模拟过程中。

f)通过简单的适应度值比较，更新每个粒子的新位置，修改 G_{best} 和 P_{best} 的值。

g)当进化过程已经达到一个满意的状态（或者最大回归数已经达到），则继续执行步骤3，否则，调整惯性权重 ω，以及模拟退火温度 $T_{i+1} = k * T_i$，设置k=k+1，并返回到步骤b)。

h)输出最好方案 G_{best} 及其适应度值。取回停止迭代时的核函数类型及参数值。

i)基于选出的核函数类型和参数，在更大训练集上反复训练和构建支持向量回归。最后，利用训练好的SVR模型在测试集上进行测试。

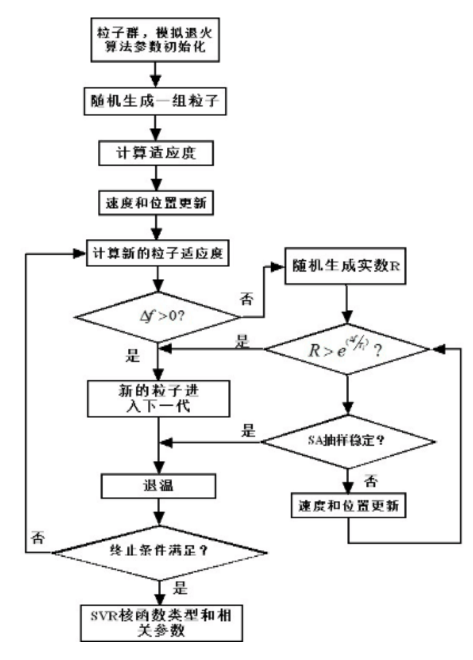


图 2　HPSOSA 优化 SVR 流程图

3　应用实例及结果分析

实验采用改进的 SVR-HPSOSA 方法，运行环境配置：英特尔赛扬 M 1.86 GHz 处理器，内存 1.5 GB，Windows XP 操作系统和 Matlab 开发环境。PSO 参数设置如下：迭代次数为 100；种群数量为 100；最小惯性权重为 0.1，最大惯性权重为 0.9，学习率为 2。初始温度为 5000；终止温度为 0.9，温度降低系数为 0.001。

3.1　实验数据

广西柳江属珠江水系，干流全长 773.3 公里，流域面积 57173 平方公里，跨黔、桂、湘三省 30 个县、市，天然落差 1306 米，平均比降 1.68‰，年均流量 1865 立方米/秒，汛期为 4-9 月。到了汛期就柳江河段而言，洪水发生较为频繁，从 1988 年至今的近 20 年间，大洪水的发生较为频繁，灾害也很严重，发生超过 100 年一遇的特大洪水 1 次（“1996.7”），20 至 50 年一遇的大洪水 3 次（“1988.8”、“1994.6”和“2004.7”洪水），10 至 20 年一遇的较大洪水 1 次（“2000.6”洪水）。为此建立较为准确的汛期柳江径流预报模型，对于防灾减灾意义十分重要。本文建立模型的数据选取广西柳江老桥口的每天 12 时的流速（m^3/s）数据进行应用实例分析，总数据为 2008 年 1 月 1 日至 2010 年 9 月 30 日期间共 1004 个数据点。将数据分为两部分，2008 年 1 月 1 日到 2010 年 3 月 31 日期间的 821 个数据点用于模型训练，余下的 183 个样本（2010 年 4 月 1 日到 2010 年 9 月 30 日为主要汛期）用于模型测试。

3.2　输入变量

一般来说，降雨-径流涉及的因果变量关系包括降雨量、降雨前水位、水流蒸发、温度等。大多数研究应用降雨量和前期流量（或水位）作为输入变量。根据逐步回归方法，同时考虑前期降雨量和径流时间序列，更能与 $X_{(t+1)}$ 相关，因此，选择这两个最佳的输入组合作为此模型的输入。由文献[15]，考虑将径流时间点 $t-1(X_{(t-1)})$、$t-2(X_{(t-2)})$、$t-3(X_{(t-3)})$、$t-4(X_{(t-4)})$ 的值及降雨量时间点 $t-1(R_{(t-1)})$ 的值作为模型的输入，通过逐步回归方法消除低影响因子并选择出最优影响力的因子。

3.3　SVR-HPSOSA 模型训练

使用相同的预测因子，对 SVR-HPSOSA 模型预测结果与 SVR 模型结果进行比较。本研究进行了简单的参数设置，设置高斯核函数带宽 δ=0.75，另外两个参数，正则常数 C 及损失函数 ε 的选择基于试验-误差法。具有测试最小均方误差平方根（RMSE）的参数被认为是最优的。得到了基于最好测试和预测结果（使 RMSE 最小）下的最优参数是（C=8，ε=0.07）。

3.4　性能评价

在许多文献中有很多关于降雨-径流预报模型性能指标。本文使用以下的评价指标来衡量该模型的性能：均方根误差（RMSE），平均绝对百分比误差（MAPE）及效率系数（CE）。此外，由于降雨径流模型峰值点非常重要，本文使

用式（12）来比较不同模型在捕捉径流时间序列的整个径流数据峰值的性能指标。

$$CE_{peak}=1-\frac{\sum_{i=1}^{n}(x_{pi}-x_{pi}^{*})^{2}}{\sum_{i=1}^{n}(x_{pi}-\bar{x}_{p})^{2}} \qquad (12)$$

CE_{peak} 是峰值的效率系数，x_{pi}，x_{pi}^{*}，$\bar{x}_{p}$ 和 n 分别代表观测值、预测值和观测平均值，以及观测到的数据峰值个数。

性能评价指标（RMSE）和（MAPE）主要是从数值上评价模型的性能，其值越小说明模型的预测性能越好；性能指标（CE）主要是看模型能否对径流的趋势做出正确的判断，其值越大就说明模型能够较为准确把握径流未来趋势；性能指标公式(12)是洪水峰值趋势的判断指标，对于径流模型在实际应用中，最关心的就是能否较为准确预测灾害性洪水，即洪水峰值达到时刻和对洪水峰值趋势跟踪，本文把流速大于 4000（m^3/s）看做是洪峰值，其中(12)中的 n 就是在预测时间区间中大于 4000（m^3/s）样本个数。

3.5 结果分析

图3为2010年4月1日到9月30日，对小于4000的径流，SVR-HPSOSA和单纯SVR（高斯核）两种模型分别得到的预测值与观测值的比较。图4是2010年4月1日到9月30日，对大于4000的径流，分别使用单纯SVR（高斯核）和提出的SVR-HPSOSA模型得到的预测峰值与观测峰值的比较。通过两图我们可以看到SVR-HPSOSA模型的结果比SVR模型更接近实际观察到的水位值。特别是，SVR-HPSOSA模型对径流峰值预测能产生更好的预测结果，而其他模型估计略高或略低。

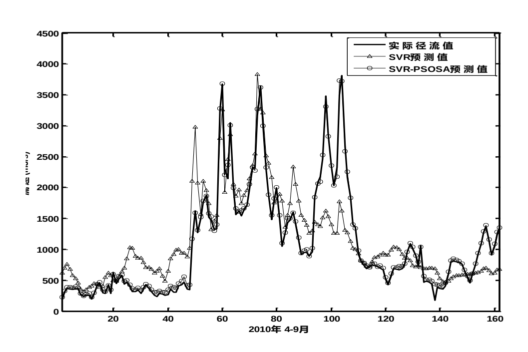


图 3 小于 4000 径流观测值与两种模型预测结果比较

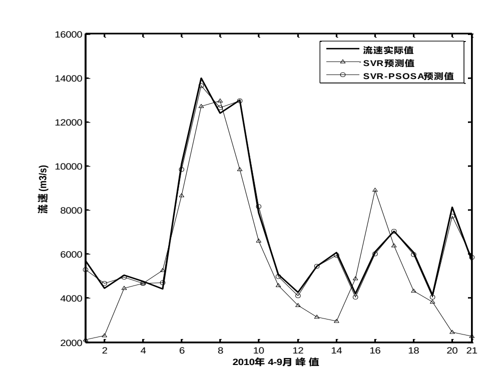


图 4 大于 4000 径流观测值与两种模型预测结果比较

表2以图表的方式分别给出了使用纯SVR（高斯核）和SVR-HPSOSA两种模型的拟合精度、预测精度和模型效率三种性能指标对比。正如表2所示，对于训练数据，SVR模型的均方根误差为167.863；而SVR-HPSOSA模型，误差只为74.218。同样，对于测试数据，SVR模型的均方根误差为909.564；而对于SVR-HPSOSA模型，误差达到201.723。针对训练和测试数据的MAPE指标，SVR-HPSOSA模型也低于SVR模型。这些结果表明，观察和预测值之间的偏差是非常小的。此外，SVR-HPSOSA模型的CE效率系数是所有模型中最大的。这也意味着SVR-HPSOSA模型能够捕获到日径流的平均变化趋势。

实验结果表明，SVR-HPSOSA模型在径流预测上是非常有前途的，既有良好的拟合表现，也有有效的预测表现。体现了SVR-HPSOSA模型训练和测试结果的一致性。从RMSE、MAPE和CE三个性能指标的表现来看，对具有相同输入的径流预测，SVR-HPSOSA模型要优于SVR模型。

与其他模型相比，SVR-HPSOSA模型对峰值预测能力。峰值估计通常是在实际降雨-径流应用中防洪减灾计划最重要参考之一，被作为建模过程考查的关键。从表2可以看出，SVR-HPSOSA模型对极端径流数据有更好预测能力，大大超过了SVR模型。SVR-HPSOSA模型训练数据的 CE_{peak} 值为0.984，测试数据是0.977。显示出了对峰值数据较高的处理能力。因此，本文提出的模型适合用于建立柳江汛期径流SVR预测模型，而且预测精度高。

从表2可以进一步看出，径流模型核函数类型最佳的选择是Sigmoid曲线核函数，最优的核参数设置为：α 等于0.325，β 为0.0264，SVR最优的参数设置：C 为18.60，ε 为0.1059。

表 2 SVR 和 SVR-HPSOSA 模型径流预测结果比较

模拟数据(2008 年 1 月 1 日- 2009 年 3 月 31 日)				
模型	MAPE	RMSE	CE	CEpeak
SVR	15.758%	167.863	0.987	0.929
SVR-HPSOSA	3.734%	74.218	0.997	0.984
预测数据(2010 年 4 月 1 日- 2010 年年 9 月 30 日)				
模型	MAPE	RMSE	CE	CEpeak
SVR	25.881%	909.564	0.822	0.482
SVR-HPSOSA	5.284%	201.723	0.992	0.977
核函数类型和参数				
	α	β	C	ε
Sigmoid	0.325	0.0264	18.60	0.1059

注：拟合模型中的径流峰值为48个样本，径流峰值为21个样本，径流值大于4000，为峰值

4 结束语

降雨-径流系统是最活跃的动态气象系统之一。本文的优势和关键在于融合PSO和SA各自的优势进化SVR核函数类型及

参数，用以建立一种径流预报模型SVR-HPSOSA。在真实降雨-径流数据实验中，训练出了SVR模型合适的核函数类型和参数值，与纯粹SVR模型预测进行比较，SVR-HPSOSA模型预测显示出了相对高的精度。根据本文的研究结果，SVR-HPSOSA方法具有以下的特点：

a）SVR-HPSOSA算法能避免局部最优解，通过混合SA的接受准则和PSO随机接受原则，减少了局部搜索端点附近的震动偏离；根据新旧位置之间不同的适应度值，运用规则接受新位置，还是重新计算另一个可信位置。这使得模型可以跳出局部最优解，并减少局部搜索终点附近的震动偏离。这些特性提高了解决方案的质量和收敛速度。

b）SVR-HPSOSA算法具有运算速度快和能在全局方向搜索最优解的优点，通过结合PSO运行机制、并行处理干扰机制以及SA解决搜索路径的特点，可以快速找到一个好的解决方案。

c）根据本文的研究结果，我们可以得出以下结论：S曲线核函数的SVR模型可以作为实际径流预测应用中的一种手段，从实证结果的角度看，它能得到较好的预报精度，具有较高的预测质量。

d）采用Sigmoid曲线核函数时，支持向量机实现的是一种前向神经网络，但是由于使用SVR方法，网络隐含层节点数、连接权值都是在设计的过程中自动确定，而且SVR的理论决定了它最终求得的是全局最优值而不是局部最小值，也保证了它对于未知样本的良好泛化能力而不会出现过学习现象，以此可以解释Sigmoid曲线核函数的SVR模型最为适宜径流预测模型。可以推广在其它地区的径流预测上，同样可以被认为是水文研究系统管理中较强的径流预测工具。

参考文献：

[1] Wu Jiansheng, Liu Mingzhe, Jin Long. A hybrid support vector regression approach for rainfall forecasting using particle swarm optimization and projection pursuit technology [J]. International Journal of Computational Intelligence and Applications, 2010, 9 (2): 87-104.

[2] 滕少华，樊继慧，陈潇等. SVM 多分类器协同挖掘局域气象数据 [J]. 广西大学学报：自然科学版, 2014, 39 (5): 1131-1137. (Teng Shaohua, Fan Jihui, Chen Xiao, etal. Application of SVM-based multi-classifiers in miningcooperatively local area meteorological data [J]. Journal of Guangxi University: Natural Science Edition, 2014, 39 (5): 1131-1137.)

[3] Wu Jiansheng. An effective hybrid semi-parametric regression strategy for rainfall forecasting combining linear and nonlinear regression [J]. International Journal of Applied Evolutionary Computation, 2011, 2 (4): 50-65.

[4] Talei A, Chua L H C, Quek C. A novel application of a neuro-fuzzy computational technique in event-based rainfall–runoff modeling [J]. Expert Systems with Applications, 2010, 37 (12): 7456-7468.

[5] 张成刚，宋佳智，姜静清，等. 一种改进的降噪自编码神经网络不平衡数据分类算法 [J]. 计算机应用研究, 2017, 34 (5): 1329-1332. (Zhang Chenggang, Song Jiazhi, Jiang Jingqing, etal. Imbalanced data classification algorithm of improved de-noising auto-encoder neural network [J]. Application Research of Computers, 2017, 34 (5): 1329-1332.)

[6] Kisi O, Shiri J, Tombul M. Modeling rainfall-runoff process using soft computing techniques [J]. Computers & Geosciences, 2013, 51 (2): 108-117.

[7] Hong Weichang, Dong Yucheng, Zheng Feifeng, *et al.* Hybrid evolutionary algorithms in a SVR traffic flow forecasting model [J]. Applied Mathematics & Computation, 2011, 217 (15): 6733-6747.

[8] 张淑娟，邓秀勤，刘波. 基于粒子群优化的最小二乘支持向量机税收预测模型研究 [J]. 计算机科学, 2017, 44 (s1): 119-122. (Zhang Shujuan, Deng Xiuqin, Liu Bo. Research on tax forecasting model based on PSO and least squares support vector machine [J]. Computer Science, 2017, 44 (s1): 119-122.)

[9] Trelea I C. The particle swarm optimization algorithm: convergence analysis and parameter selection [J]. Information Processing Letters, 2016, 85 (6): 317-325.

[10] Chakrapani B, Varadarajan V. Genetic algorithm based demand side management for smart grid [J]. Wireless Personal Communications, 2017, 93 (2): 481-502.

[11] 李学俊，徐佳，朱二周，等. 任务调度算法中新的自适应惯性权重计算方法 [J]. 计算机研究与发展, 2016, 53 (9): 1990-1999. (Li Xuejun, Xu Jia, Zhu Erzhou, etal. A novel computation method for adaptive inertia weight of task scheduling algorithm [J]. Journal of Computer Research and Development, 2016, 53 (9): 1990-1999.)

[12] Ireneusz G. Discussion on semi-immune algorithm behaviour based on fractal analysis [J]. Soft Computing, 2017, 21 (14): 3945-3956.

[13] Bermolen P, Rossi D. Support vector regression for link load prediction [J]. Computer Networks, 2009, 53 (2): 191-201.

[14] 吴意乐，何庆. 基于改进遗传模拟退火算法的 WSN 路径优化算法 [J]. 计算机应用研究, 2016, 33 (10): 2959-2962. (Wu Yile, He Qing. WSN path optimization algorithm based on improved genetic simulated annealing algorithm [J]. Application Research of Computers, 2016, 33 (10): 2959-2962.)

[15] Asadi S, Shahrabi J, Abbaszadeh P, *et al.* A new hybrid artificial neural networks for rainfall-runoff process modeling [J]. Neurocomputing, 2013, 121 (18): 470-480.